

Jornadas de Automática

Localización en viñedos mediante LiDAR 3D y arquitectura neuronal multitarea

Judith Vilella-Cantos^{a,*}, Míriam Máximo^a, Óscar Reinoso^a, Arturo Gil^a, Mónica Ballesta^a, David Valiente^a

^aInstituto de Investigación en Ingeniería de Elche (I3E), Universidad Miguel Hernández de Elche, Avda. de la Universidad s/n, Edificio Innova, 03202, Elche, Alicante, España.

To cite this article: Vilella-Cantos, J., Máximo, M., Reinoso, O., Gil, A., Ballesta, M., Valiente, D. 2026. Localiza-tion in vineyards through 3D LiDAR and multi-task neural architecture. Jornadas de Automática, 47. <https://doi.org/10.17979/ja-cea.2026.47.13543>

Resumen

Obtener una localización precisa en un entorno agrícola es un problema complejo que todavía no ha encontrado una solución robusta en el estado del arte. Por ejemplo, en un viñedo, la repetición de las hileras dificulta la identificación de elementos característicos que permitan diferenciar los lugares para una tarea de reconocimiento de lugares. En este trabajo, nos centramos en la tarea de reconocimiento de lugares con sensores LiDAR en entornos de viñedo. Hemos ampliado una red neuronal convolucional para que realice una tarea adicional: clasificar el lugar en función de etiquetas que indican si la nube de puntos pertenece a una zona de giro del robot o en el interior de una hilera. Proporcionamos un conjunto de entrenamiento que cubre distintas etapas de crecimiento de la vid, mediante los datos del dataset TEMPO-VINE. Los resultados demuestran que, con una pequeña modificación, la misma arquitectura diseñada para el reconocimiento de lugares con LiDAR puede predecir con una precisión superior al 85 % la etiqueta del lugar al que pertenece. Estos avances allanan el camino para mejorar las soluciones existentes de reconocimiento de lugares con LiDAR en entornos agrícolas, en creciente expansión, mediante la extracción de información adicional con la misma arquitectura.

Palabras clave: Robótica móvil, Reconocimiento de patrones e IA en agricultura, Robótica agrícola, Aprendizaje automático, Localización, Robots móviles autónomos

Localization in vineyards through 3D LiDAR and multi-task neural architecture

Abstract

Obtaining precise location data in an agricultural setting is a complex problem for which no robust solution has yet been found using state-of-the-art methods. For example, in a vineyard, the repetition of rows makes it difficult to identify distinctive features that would allow for the differentiation of locations in a place recognition task. In this work, we focus on the task of place recognition in vineyard environments using LiDAR sensors. We have extended a convolutional neural network to perform an additional task: classifying the location based on labels that indicate whether the point cloud belongs to a robot turning zone or within a row. We provide a training dataset covering different stages of vine growth, using data from the TEMPO-VINE dataset. The results show that, with a minor modification, the same architecture designed for LiDAR-based location recognition can predict the location label with over 85 % precision. These advances pave the way for improving existing LiDAR-based place recognition solutions in agricultural settings, which are growing rapidly, by extracting additional information using the same architecture.

Keywords: Mobile robotics, Pattern recognition and AI in agriculture, Agricultural robotics, Machine Learning, Localization, Autonomous Mobile Robots

*Autor para correspondencia: jvilella@umh.es
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

1. Introducción

En entornos rurales, encontrar una solución robusta al problema de *place recognition* (reconocimiento de lugares) es una tarea compleja, debido a la naturaleza repetitiva de estos entornos. El hecho de encontrar una solución fiable a este problema tendría un impacto altamente beneficioso en nuestra sociedad, principalmente debido al aumento progresivo de la población mundial y la falta de trabajadores agrarios, estando la media de edad de estos últimos cada vez más elevada (Ali et al., 2025; Eze et al., 2025).

Obtener una solución robusta al problema de la localización de robots móviles en estos entornos supone muchos beneficios para la industria agrónoma, como por ejemplo la operabilidad entre hileras del cultivo a la hora de monitorizar su estado o realizar tareas de fumigado o recolección (Zheng et al., 2024). En entornos urbanos, el problema del reconocimiento de lugares basado en LiDAR (*LiDAR place recognition*) ha sido estudiado en detalle en los últimos diez años, alcanzando resultados excepcionales (Xia et al., 2023; Joseph et al., 2025). No obstante, obtener resultados igual de robustos en entornos desestructurados es todavía una tarea pendiente (Sun et al., 2024; Maweni and Amayo, 2025).

El uso de etiquetas ha demostrado ser de gran valor a la hora de realizar tareas de visión por computador en estos entornos, como el monitorizado de los cultivos o el conteo de frutos en huertos (Lobefaro et al., 2025). Trabajos recientes también evidencian que la información semántica puede usarse para obtener soluciones robustas de navegación autónoma en entornos desestructurados (De Silva et al., 2025; Audrito et al., 2026).

En este trabajo, presentamos un método que simultáneamente ofrece una solución al problema de *LiDAR place recognition*, y, a su vez, de clasificación de cada nube de puntos en función de la zona del cultivo a la pertenece, usando la misma estructura de red neuronal de tipo codificador-decodificador. Hemos validado nuestro método en un dataset de viñedos y ofrecemos resultados para distintas etapas de crecimiento del fruto, evidenciando el reto que supone el hecho de encontrar una solución robusta en este contexto.

El resto del artículo se organiza de la siguiente manera: la Sección 2 presenta el estado del arte del tema de estudio. La Sección 3 describe la metodología que hemos adoptado para desarrollar una red neuronal multitarea y etiquetar los datos. La Sección 4 resume los resultados obtenidos por nuestro método para distintos meses individuales y etapas fenológicas de un viñedo. Finalmente, la Sección 5 presenta las conclusiones obtenidas del método propuesto, así como futuras líneas de mejora.

2. Estado del arte

La automatización de tareas en entornos agrícolas, especialmente en cultivos de alto valor como los viñedos, ha impulsado el desarrollo de vehículos autónomos capaces de navegar de forma precisa. Sin embargo, la localización robusta sigue siendo un desafío abierto. A diferencia de los entornos urbanos, los campos agrícolas presentan una alta ambigüedad perceptual y una carencia de puntos de referencia distintivos.

En particular, la estructura geométrica de los viñedos se caracteriza por hileras paralelas y repetitivas, lo que dificulta las tareas de *place recognition* debido a la similitud entre diferentes secciones del cultivo (Salveti et al., 2022).

El uso de sensores LiDAR ha ganado tracción en la robótica de exteriores gracias a su resiliencia frente a cambios de iluminación. No obstante, las soluciones convencionales basadas en nubes de puntos suelen centrarse en la extracción de descriptores globales o locales que, en entornos de hileras, resultan ser muy similares entre sí, provocando falsos positivos en la detección de bucles (*loop closure*). Soluciones recientes como ORCHNet (Barros et al., 2023) o PointNetPGAP-SLC (Barros et al., 2024) enfocan específicamente a entornos agrícolas el problema de *LiDAR place recognition* mediante enfoques novedosos que se basan en la agregación de características y el uso de la información de cada hilera en la función de pérdida para distinguir entre positivos y negativos. No obstante, este tipo de soluciones no alcanzan la robustez de métodos diseñados y probados en entornos estructurados como MinkUNeXt (Cabrera et al., 2025; Vilella-Cantos et al., 2025) o MinkLoc3D (Komorowski, 2021), que alcanzan un *recall* superior al 90 %.

Investigaciones recientes sugieren que la integración de tareas secundarias en el entrenamiento de redes neuronales puede mejorar la robustez de la tarea principal, o al menos complementarla obteniendo resultados competitivos en un campo auxiliar. Este enfoque se le conoce como aprendizaje multitarea y se ha aplicado en diferentes ámbitos, desde el transporte inteligente (Alzahrani et al., 2024), el diseño de medicamentos (Allenspach et al., 2024) hasta el procesamiento del lenguaje natural (Chen et al., 2024). Estudios como el de Lin et al. (Lin et al., 2023) evidencian que la ponderación entre las diferentes tareas es crucial en este tipo de sistemas.

En este contexto, el uso de datasets multitemporales, como TEMPO-VINE (Martini et al., 2025) o Bacchus Long-Term (BLT) (Polvara et al., 2024), permite abordar la variabilidad fenológica de la vid (cambios en el follaje según la estación), un factor que suele degradar el rendimiento de los algoritmos de localización a largo plazo. La tendencia actual se dirige hacia modelos que no solo identifiquen "dónde" está el robot, sino que comprendan el contexto topológico del entorno (por ejemplo, si se encuentra en zona de maniobra o de trabajo lineal), permitiendo así filtrar candidatos y reducir el error en la localización global.

3. Metodología

Para alcanzar nuestro propósito, hemos tomado dos pasos: primero, el etiquetado de los datos del dataset TEMPO-VINE (Martini et al., 2025), de tal modo que se tenga una etiqueta de zona para cada nube de puntos. Una vez etiquetados los datos, esta información se pasa como entrada a nuestro método propuesto, el cual consiste en una modificación del método MinkUNeXt-VINE (Vilella-Cantos et al., 2026) al añadir capas para realizar una tarea adicional de predicción de zona. En los siguientes subapartados proporcionamos más información al respecto de ambas contribuciones.

3.1. Etiquetado del viñedo

El dataset TEMPO-VINE ofrece datos capturados durante todo el año 2025 (desde febrero hasta noviembre) para dos tipos de viñedo: espaldera (descubierto) y pérgola (la vegetación crea una cubierta sobre el campo). La Figura 1 resalta la diferencia entre estos dos tipos de cultivo.



Figura 1: Los dos tipos de viñedo capturados por el dataset TEMPO-VINE. A la izquierda: espaldera. A la derecha: pérgola.

En este trabajo, nos centramos en etiquetar los datos del viñedo de tipo espaldera, empleando tres tipos de etiqueta distintos. El valor “0” se corresponde con las nubes de puntos situadas en el extremo inicial del viñedo; el valor “1”, con las nubes de puntos situadas en el extremo final del viñedo. Finalmente, el valor “2” se refiere a las nubes de puntos situadas en el interior de cualquiera de las hileras del viñedo.

Para llevar a cabo el etiquetado de manera automática, se ha desarrollado un código Python para este propósito. El código en cuestión carga los datos del GPS proyectados en metros (UTM) de cada ruta capturada por el dataset y los alinea mediante el análisis de componentes principales (PCA). Este proceso normaliza la geometría del viñedo y facilita el análisis espacial. Dado que la mayor varianza de los datos GPS coincide con la trayectoria longitudinal de las cepas, el PCA permite reorientar el sistema de coordenadas de modo que el eje principal se alinee con las filas y el secundario con el eje transversal. Esta transformación matemática permite segmentar de forma robusta las zonas de trabajo (interior de las hileras) de las áreas de maniobra (cabeceras) mediante un filtrado por percentiles en el nuevo eje longitudinal.

La Figura 2 muestra una imagen satelital del viñedo donde el dataset TEMPO-VINE fue capturado, con las posiciones GPS etiquetadas por zonas siguiendo el procedimiento previamente mencionado.

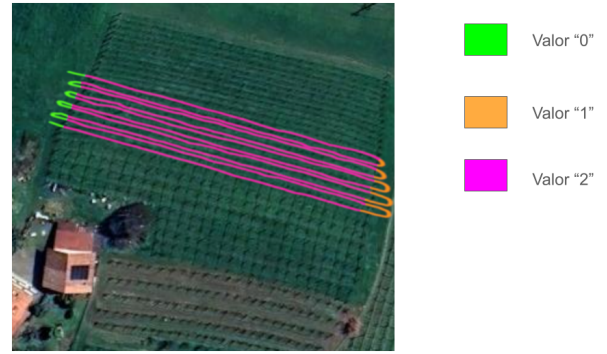


Figura 2: Vista satelital del viñedo sobre el cual el dataset TEMPO-VINE fue capturado, incluyendo las etiquetas por zona calculadas. El viñedo está localizado en la zona norte de Italia (Agliè, Turín).

3.2. Modificación de la arquitectura

Para conseguir que una red inicialmente diseñada para realizar exclusivamente la tarea de *LiDAR place recognition* también realice una tarea auxiliar, nos hemos apoyado en la inclusión de capas lineales. Debido a que MinkUNeXt-VINE se trata de una red de tipo codificador-decodificador, hemos situado estas nuevas capas en la salida del codificador, antes de que las siguientes capas de la red realicen la reconstrucción de la nube de puntos de entrada a partir de su versión codificada. Las capas nuevas en cuestión consisten en una capa lineal que reduce la dimensionalidad de los datos a la mitad, una capa de normalización y una RELU para estabilizar el proceso, finalizando con una capa lineal que pasa la dimensionalidad a tres (las tres clases que la red debe aprender a distinguir). La Figura 3 muestra la arquitectura final propuesta.

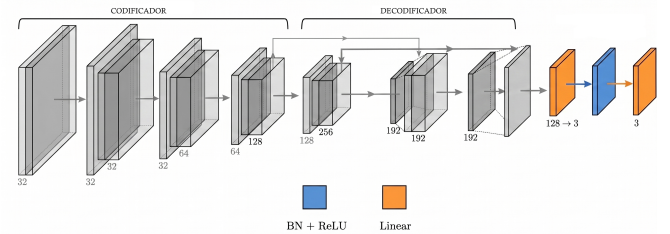


Figura 3: Arquitectura del método multi-tarea propuesto.

En cuanto a la función de pérdida de nuestro método, es importante tener en cuenta que nuestra nueva tarea auxiliar de clasificación de zona necesita su propia función de pérdida. La función de pérdida del método propuesto se trata de una suma ponderada, de tal modo que:

$$\mathcal{P}_{\text{total}} = \mathcal{P}_{\text{LPR}} + \alpha \mathcal{P}_{\text{clas}} \quad (1)$$

Siendo $\mathcal{P}_{\text{total}}$ la función de pérdida de nuestro método, $\mathcal{P}_{\text{LPR}}$ la función de pérdida para la tarea de *LiDAR place recognition* usada en la arquitectura original, α el peso asignado a la tarea auxiliar. En nuestro sistema, el valor de α se ha establecido en 0,5. Este valor se ha elegido después de realizar distintos experimentos para diversos valores de $\alpha \in [0, 0,25, 0,50, 0,75, 1,0]$ bajo las mismas condiciones: datos del viñedo capturados en

Febrero, con una trayectoria como conjunto de entrenamiento y la otra como conjunto de validación. La estructura del dataset utilizado se detalla en la Sección 4. La Tabla 1 muestra los resultados para estas evaluaciones.

α	Recall@1 %	Recall@1	Accuracy	F1-Score
0	0.9730	0.6777	0	0
0.25	0.9596	0.6360	0.9322	0.9389
0.5	0.9632	0.6458	0.9452	0.9498
0.75	0.9583	0.6164	0.9235	0.9317
1.0	0.9608	0.6188	0.9452	0.9497

Tabla 1: Resultados para ambas tareas con nuestro método y distintos valores de α .

Un valor de α de 0,5 permite que la red se beneficie de la información complementaria de la tarea auxiliar sin que esta interfiera en la convergencia de la tarea principal. $\mathcal{P}_{\text{clas}}$ representa la función de pérdida que usamos para la tarea de clasificación por zonas. La definición de la función de pérdida para esta tarea adicional es la siguiente:

$$\mathcal{P}_{\text{clas}} = - \sum_{i=1}^C \alpha_i y_i \log(p_i) = -\alpha_y \log(p_y), \quad (2)$$

en este contexto, C es el total de clases en nuestro sistema (en nuestro caso, tres), α_i el peso asociado a la clase i , y_i es la etiqueta de clase correcta para el ground truth. La segunda igualdad de la ecuación 2 representa la versión simplificada, teniendo en cuenta que el valor de y_i será 1 si la predicción es correcta y 0 en caso contrario. Finalmente, p_i representa la probabilidad de predicción de la clase i , calculada mediante una función *softmax* aplicada a la salida de nuestro nuevo cabezal. Se trata de una función de pérdida de tipo *cross entropy* ponderada (Aurelio et al., 2019), la cual funciona muy bien en nuestro contexto, ya que tenemos muchas más muestras de una clase (interior de la hilera del viñedo) que del resto.

En cuanto a la selección de los pesos α_c asociados a cada etiqueta de clase c , se ha adoptado una estrategia de balanceado por frecuencia media, siguiendo la siguiente fórmula:

$$\alpha_c = \frac{f_{\text{med}}}{f_c} \quad (3)$$

donde f_{med} representa la frecuencia media para cada clase presente en el conjunto, mientras que f_c representa la frecuencia de la clase dada c . De este modo, nos aseguramos que la ponderación de cada una de las clases sea coherente con el número de muestras presentes para cada una de ellas, reduciendo el impacto del desbalanceo de datos para las clases correspondientes a los extremos del viñedo.

4. Resultados y discusión

En esta sección, proporcionamos resultados cuantitativos de nuestro método en dos escenarios: usando campañas individuales para entrenar y validar nuestra red, o bien usando paquetes de varias campañas representativos de distintas etapas de crecimiento de la cosecha. Las nubes de puntos fueron capturadas por un sensor Velodyne VLP-16. Cada campaña consta de dos trayectorias: “run1”, la cual consiste en 10 hileras del

viñedo descubierto tomadas de manera consecutiva; y “run2”, que graba 8 hileras, la mitad de ellas grabadas en sentido inverso al caso de “run1”. En estos experimentos, dividimos las zonas de test a partir de posiciones GPS y un radio asociado a dicha posición (23 metros). Todas las nubes que pertenezcan a dicha zona se consideran el conjunto de test, siguiendo la metodología propuesta en (Uy and Lee, 2018).

Para evaluar la tarea principal de localización, las métricas utilizadas se basan en el *Recall*, el cual representa cuántas de las predicciones de lugares de la red realmente se encontraban en el conjunto de posiciones correctas. Concretamente, se reportan las métricas *Recall@1* y *Recall@1 %*, las cuales miden el porcentaje de consultas en las que se encuentra al menos una coincidencia correcta dentro del primer puesto y del primer 1 % de los resultados recuperados, respectivamente. Por otro lado, la tarea de clasificación se evalúa mediante la métrica *Accuracy*, que cuantifica la proporción de aciertos totales sobre el conjunto de datos, y el *F1-score*, que ofrece una medida balanceada al calcular la media armónica entre precisión y exhaustividad.

La Tabla 2 muestra los resultados de nuestro método tanto para la tarea principal (*LiDAR place recognition*) como la auxiliar (clasificación de zona) para distintas campañas del dataset TEMPO-VINE.

En estos resultados, podemos observar que la tarea de clasificación obtiene hasta un 90 % de F1-Score para etapas similares del cultivo (septiembre-julio). No obstante, para etapas muy dispares, como febrero-septiembre, el valor para esta métrica cae hasta un 20 %. Esto evidencia que son necesarios nuevos esfuerzos para adaptar la arquitectura base para un desempeño más robusto de esta tarea. En cuanto a la tarea principal de localización, las oscilaciones en las métricas son menos agresivas, con alrededor de 80 % de Recall@1 % en el caso de usar el mes de febrero como test. En este mes, la tarea de localización es más sencilla, puesto que la falta de vegetación en el viñedo permite la visibilidad de las filas adyacentes, facilitando un mayor encuentro de marcas significativas en los escaneos. El escenario de desarrollo completo de la vegetación (septiembre) hace que resolver la tarea principal sea más complejo, con valores de Recall@1 que caen entre un 30 % y un 40 %.

Por otro lado, la Tabla 3 muestra los resultados cuantitativos en el caso en el que los meses se agrupan en paquetes, tanto en la fase de entrenamiento como de evaluación de nuestra red. El estado “Temprano” recoge los datos de las campañas de febrero, marzo y abril, donde el viñedo tiene entre ninguna y poca vegetación. El estado “Medio” recoge los meses de mayo y junio (teniendo este último dos campañas), en los que la vegetación del viñedo es abundante, dándose casos incluso en los que la hierba del suelo es alta, interfiriendo con la captura de los datos del sensor. Finalmente, el conjunto “Tardío” recoge las campañas de julio, agosto y septiembre, en las que la uva está además completamente desarrollada, acercándose la fecha de la vendimia. La Figura 4 muestra el estado de la cosecha en cada uno de los paquetes establecidos.

Entrenamiento	Test	Localización		Clasificación	
		Recall@1 %	Recall@1	Accuracy	F1-Score
Feb.	Febrero	0.8269	0.5327	0.8496	0.8389
	Julio	0.5658	0.2893	0.3118	0.3775
	Septiembre	0.6287	0.3166	0.1746	0.2103
Julio	Febrero	0.7931	0.4998	0.5338	0.6202
	Julio	0.7229	0.4361	0.7970	0.8155
	Septiembre	0.6644	0.3879	0.5661	0.6486
Sept.	Febrero	0.7683	0.4724	0.1464	0.1826
	Julio	0.6698	0.3963	0.8931	0.9047
	Septiembre	0.6772	0.4329	0.8449	0.8657

Tabla 2: Evaluación cruzada de nuestro método por meses individuales.



Figura 4: Estado del viñedo en cada uno de los distintos estados de crecimiento del cultivo. La imagen de la izquierda muestra el viñedo en estado "Temprano"; la central muestra el estado "Medio"; la de la derecha muestra el estado "Tardío", con el fruto listo para la cosecha.

Los resultados presentados en la Tabla 3 demuestran que los mejores resultados se dan al entrenar con datos capturados en etapas tempranas de la cosecha. Esto se debe a que, como se comentó en el caso de la tabla anterior, una etapa en la que la vegetación es escasa favorece la visibilidad de las zonas adyacentes y, por tanto, facilita la extracción de características. Podemos apreciar que, por el hecho de haber entrenado con varias campañas en lugar de una, la oscilación entre las métricas es menor, con resultados superando el 80 % de *Accuracy* para la tarea auxiliar en la mayoría de los casos. No obstante, al haber entrenado con datos del fruto en su estado totalmente desarrollado, obtenemos resultados degradados cuando evaluamos con datos del viñedo en invierno. Esto es debido a la disparidad marcada en la apariencia por estaciones. En cuanto a la tarea principal, cuando trabajamos con el estado temprano del cultivo exclusivamente se alcanza un 84.43 % de Recall@1 %. No obstante, el Recall@1 no alcanza el 50 % en los escenarios planteados. La explicación a este fenómeno es que, desde un punto de vista computacional, al incluir más meses en los conjuntos de validación, estamos poblando el espacio de búsqueda con una mayor variedad de descriptores. Esto reduce la "distancia de seguridad" entre las representaciones de lugares distintos pero visualmente similares bajo ciertas condiciones climáticas o de poda. En resumen, el modelo tiene más oportunidades de confundirse debido a que el espectro de apariencias posibles de las hileras se vuelve continuo y denso.

5. Conclusión y trabajo futuro

En este trabajo, presentamos una mejora sobre una red neuronal de *LiDAR place recognition* para que resuelva a su vez una tarea auxiliar: clasificación de cada nube de puntos según la zona del viñedo a la que pertenece. Para ello, hemos etiquetado los datos del *dataset* TEMPO-VINE (Martini et al., 2025) diferenciando tres zonas cruciales en un viñedo: los dos extremos y el interior de la hilera. Las modificaciones de la red consisten en un nuevo cabezal con operaciones lineales situado en la salida del codificador de la red original. Junto con una función de pérdida de tipo *cross entropy* ponderada para compensar el desbalance entre clases, nuestra propuesta consigue obtener resultados de más de 85 % de *Accuracy* para la nueva tarea auxiliar. Con la inclusión de la tarea de clasificación, el rendimiento de la tarea principal no decae, manteniéndose en los mismos valores que en la estructura original (Vilella-Cantos et al., 2026), con alrededor de un 80 % de Recall@1 % para el caso del entrenamiento con datos del viñedo en estado "Temprano".

En un futuro, nuestro objetivo es, por un lado, estabilizar el desempeño de la tarea de clasificación. Como se ha podido observar en los resultados presentados en la Sección 4, los valores en el caso de las métricas referidas a esta nueva tarea oscilan muy bruscamente dependiendo de la similitud del conjunto de datos de test con el de entrenamiento. Por otro lado, también trabajaremos en hacer que la tarea auxiliar se pueda utilizar para obtener mejores resultados en la principal, mediante nuevos cambios estructurales como la introducción de mecanismos de atención.

Agradecimientos

Esta investigación forma parte del proyecto PID2023-149575OB-I00, financiado por MICIU/AEI/10.13039/501100011033 y por FEDER, UE. También forma parte del proyecto CIPROM/2024/8, financiado por la Generalitat Valenciana, Conselleria de Educación, Cultura, Universidades y Empleo (programa PROMETEO 2025).

Referencias

Ali, A., Hussain, T., Zahid, A., 2025. Smart irrigation technologies and prospects for enhancing water use efficiency for sustainable agriculture.

Entrenamiento	Test	Localización		Clasificación	
		Recall@1 %	Recall@1	Accuracy	F1-Score
Temprano	Temprano	0.8443	0.4878	0.8844	0.8561
Temprano	Medio	0.4982	0.1676	0.8332	0.8371
Temprano	Tardío	0.5242	0.2065	0.6902	0.7356
Medio	Temprano	0.6855	0.3076	0.8789	0.8586
Medio	Medio	0.7370	0.3973	0.8549	0.8520
Medio	Tardío	0.6389	0.3254	0.8662	0.8634
Tardío	Temprano	0.7709	0.3228	0.1720	0.2005
Tardío	Medio	0.6043	0.2434	0.8551	0.8646
Tardío	Tardío	0.7097	0.3164	0.8366	0.8632

Tabla 3: Evaluación entre distintas etapas de crecimiento del fruto por paquetes de campañas.

- AgriEngineering 7 (4), 106.
DOI: 10.3390/agriengineering7040106
- Allenspach, S., Hiss, J. A., Schneider, G., 2024. Neural multi-task learning in drug design. *Nature Machine Intelligence* 6 (2), 124–137.
DOI: 10.1038/s42256-023-00785-4
- Alzahrani, M., Wang, Q., Liao, W., Chen, X., Yu, W., 2024. Survey on multi-task learning in smart transportation. *IEEE Access* 12, 17023–17044.
DOI: 10.1109/ACCESS.2024.3355034
- Audrito, G., Martini, M., Navone, A., Galluzzo, G., Chiaberge, M., 2026. VinePT-Map: Pole-Trunk Semantic Mapping for Resilient Autonomous Robotics in Vineyards. *arXiv preprint arXiv:2603.05070*.
DOI: 10.48550/arXiv.2603.05070
- Aurelio, Y. S., De Almeida, G. M., de Castro, C. L., Braga, A. P., 2019. Learning from imbalanced data sets with weighted cross-entropy function. *Neural processing letters* 50 (2), 1937–1949.
DOI: 10.1007/s11063-018-09977-1
- Barros, T., Garrote, L., Conde, P., Coombes, M., Liu, C., Premebida, C., Nunes, U., 2024. PointNetPGAP-SLC: A 3D LiDAR-Based Place Recognition Approach With Segment-Level Consistency Training for Mobile Robots in Horticulture. *IEEE Robotics and Automation Letters*.
DOI: 10.1109/LRA.2024.3475044
- Barros, T., Garrote, L., Conde, P., Coombes, M., Liu, C., Premebida, C., Nunes, U. J., 2023. ORCHNet: A robust global feature aggregation approach for 3d lidar-based place recognition in orchards. *arXiv preprint arXiv:2303.00477*.
DOI: 10.48550/arXiv.2303.00477
- Cabrera, J. J., Santo, A., Gil, A., Viegas, C., Payá, L., 2025. MinkUNeXt: Point cloud-based large-scale place recognition using 3D sparse convolutions. *Array* 28, 100569.
DOI: 10.1016/j.array.2025.100569
- Chen, S., Zhang, Y., Yang, Q., 2024. Multi-task learning in natural language processing: An overview. *ACM Computing Surveys* 56 (12), 1–32.
DOI: 10.1145/3663363
- De Silva, R., Swindell, J., Cox, J., Popović, M., Cadena, C., Stachniss, C., Polvara, R., 2025. Keypoint semantic integration for improved feature matching in outdoor agricultural environments. *IEEE Robotics and Automation Letters* 10 (12), 13383–13390.
DOI: 10.1109/LRA.2025.3629991
- Eze, V. H. U., Eze, E. C., Alaneme, G. U., BUBU, P. E., Nnadi, E. O. E., Okon, M. B., 2025. Integrating IoT sensors and machine learning for sustainable precision agroecology: enhancing crop resilience and resource efficiency through data-driven strategies, challenges, and future prospects. *Discover Agriculture* 3 (1), 83.
DOI: 10.1007/s44279-025-00247-y
- Joseph, T., Fischer, T., Milford, M., 2025. Matched Filtering Based LiDAR Place Recognition for Urban and Natural Environments. *IEEE Robotics and Automation Letters*.
DOI: 10.1109/LRA.2025.3533966
- Komorowski, J., 2021. MinkLoc3D: Point cloud based large-scale place recognition. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. pp. 1790–1799.
- Lin, B., Jiang, W., Ye, F., Zhang, Y., Chen, P., Chen, Y.-C., Liu, S., Kwok, J. T., 2023. Dual-balancing for multi-task learning. *arXiv preprint arXiv:2308.12029*.
DOI: 10.48550/arXiv.2308.12029
- Lobefaro, L., Sodano, M., Fusaro, D., Magistri, F., Malladi, M. V., Guadagnino, T., Pretto, A., Stachniss, C., 2025. Spatio-temporal consistent semantic mapping for robotics fruit growth monitoring. *IEEE Robotics and Automation Letters*.
DOI: 10.1109/LRA.2025.3594985
- Martini, M., Ambrosio, M., Vilella-Cantos, J., Navone, A., Chiaberge, M., 2025. TEMPO-VINE: A Multi-Temporal Sensor Fusion Dataset for Localization and Mapping in Vineyards. *arXiv preprint arXiv:2512.04772*.
DOI: 10.48550/arXiv.2512.04772
- Maweni, T., Amayo, P., 2025. Adapting SegMap: A LiDAR place recognition framework for standalone use in C++ applications. In: *MATEC Web of Conferences*. Vol. 417. EDP Sciences, p. 04004.
DOI: 10.1051/mateconf/202541704004
- Polvara, R., Molina, S., Hroob, I., Papadimitriou, A., Tsiolis, K., Giakoumis, D., Likothanassis, S., Tzovaras, D., Cielniak, G., Hanheide, M., 2024. Bacchus Long-Term (BLT) data set: Acquisition of the agricultural multimodal BLT data set with automated robot deployment. *Journal of Field Robotics* 41 (7), 2280–2298.
DOI: 10.1002/rob.22228
- Salveti, F., Angarano, S., Martini, M., Cerrato, S., Chiaberge, M., 2022. Waypoint generation in row-based crops with deep learning and contrastive clustering. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 203–218.
DOI: 10.1007/978-3-031-26422-1_3
- Sun, N., Fan, Z., Qiu, Q., Li, T., Feng, Q., Ji, C., Zhao, C., 2024. TriLoc-NetVLAD: Enhancing Long-term Place Recognition in Orchards with a Novel LiDAR-Based Approach. In: *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 16–22.
DOI: 10.1109/IROS58592.2024.10802261
- Uy, M. A., Lee, G. H., 2018. PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4470–4479.
- Vilella-Cantos, J., Cabrera, J. J., Payá, L., Ballesta, M., Valiente, D., 2025. MinkUNeXt-SI: Improving point cloud-based place recognition including spherical coordinates and LiDAR intensity. *arXiv preprint arXiv:2505.17591*.
DOI: 10.48550/arXiv.2505.17591
- Vilella-Cantos, J., Martini, M., Chiaberge, M., Ballesta, M., Valiente, D., 2026. Low cost, high efficiency: Lidar place recognition in vineyards with matryoshka representation learning. *Ecological Informatics* 95, 103780.
DOI: 10.1016/j.ecoinf.2026.103780
- Xia, Y., Gladkova, M., Wang, R., Li, Q., Stilla, U., Henriques, J. F., Cremers, D., 2023. CASSPR: Cross attention single scan place recognition. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 8461–8472.
- Zheng, Z., Chen, L., Wei, L., Huang, W., Du, D., Qin, G., Yang, Z., Wang, S., 2024. An efficient and lightweight banana detection and localization system based on deep cnns for agricultural robots. *Smart Agricultural Technology* 9, 100550.
DOI: 10.1016/j.atech.2024.100550